QAPNet: A Quantum-Attentive Patchwise Network for Robust Medical Image Classification Under Noisy Inputs

Maqsudur Rahman[1,2], **Jun Zhuang**[1]

[1]Boise State University, Boise, ID 837206, USA
[2]Comilla University, Cumilla 3500, Bangladesh
mrrajon@cou.ac.bd, junzhuang@boisestate.edu

Abstract

Robust medical image classification under input corruption and bag-level annotation remains a critical challenge in clinical AI applications. We propose **QAPNet**, a Quantum-Attentive Patchwise Network that integrates quantum neural encoding, additive attention-based instance reweighting, and prototype-contrastive regularization for reliable diagnosis from degraded inputs. Our framework uses a sliding-window strategy to divide each MRI medical Image into overlapping patches, where each is encoded via an 8-qubit quantum circuit using RY-based noise-sensitive layers for yielding expressive low-dimensional representations without relying on classical CNNs. A lightweight additive attention mechanism computes instance-wise importance weights that enable interpretable and noise-aware bag-level aggregation. To enhance robustness, we apply a contrastive loss that aligns clean and noisy embeddings and enforce prototype-guided clustering via class-wise centroids. We evaluate QAPNet across seven benchmark medical imaging datasets under three levels of additive Gaussian noise ($\sigma \in \{5\%, 10\%, 30\%\}$). QAPNet consistently outperforms eight strong baselines and achieves up to $+20.8\%$ higher accuracy in OASIS (with 30% noise), $+17.7\%$ in PathMNIST, and maintains stable performance ($< 4\%$ degradation) in all settings. Ablation studies confirm the critical role of quantum encoding, attention-based aggregation, and prototype contrastive learning. These results suggest that QAPNet offers a scalable and interpretable architecture for noisy medical imaging tasks in the real world to bridge the quantum representation learning with robust clinical prediction.

Code — https://github.com/mrrajon04/QAPNet.

Introduction

Medical image classification in clinical settings often suffers from input-level corruption where image quality is degraded by acquisition noise such as Gaussian noise, scanner variability, or modality-specific artifacts to obscuring critical diagnostic patterns and hindering model generalization (Karimi et al. 2020). These challenges are particularly problematic for classical deep learning models, which tend to overfit noisy inputs and require large-scale annotated datasets that are scarce in medical domains (Xue et al. 2022a). Quantum Neural Networks (QNNs) built on variational quantum circuits (VQCs) (Zhuang, Cunningham, and Guan 2024) provide a promising low-parameter alternative to learn expressive nonlinear features in low-sample and high-dimensional regimes (Mathur et al. 2021; Liu et al. 2021). Prior work has demonstrated the superiority of QNNs applied in various areas, such as mitigating barren plateaus (Cunningham and Zhuang 2025; Zhuang and Guan 2025), medical imaging (Mazher et al. 2024a; Bokhan et al. 2022; Yousif, Al-Khateeb, and Garcia-Zapirain 2024a), and hybrid-quantum transfer learning (Dhara, Agrawal, and Roy 2024). However, these models are designed for clean inputs and degrade under image noise, as they lack mechanisms to model patch-level uncertainty or enforce structured semantic alignment. Our re-implementation confirms significant accuracy drops under Gaussian corruption, revealing a critical limitation in current quantum learning frameworks.

To address these limitations, we propose **QAPNet**, a Quantum-Attentive Patch-level Network for robust medical image classification under noisy image inputs and bag-level annotation. Unlike prior quantum methods that depend on handcrafted patching or classical CNN extractors, QAPNet introduces a fully hybrid architecture integrating: (i) **patch-level quantum encoding** via an 8-qubit variational quantum circuit (VQC) for low-parameter non-classical feature representation; (ii) a novel **element-wise additive attention** mechanism for per-instance reweighting without dot-product interactions for improving noise robustness; (iii) **contrastive learning** between clean and Gaussian-noised views to enforce embedding stability; and (iv) **prototype-guided regularization** to anchor embeddings to class centroids for promoting semantic structure and interpretability under corruption. Our **motivation** stems from empirical observations: existing QNN-based models degrade under image corruption due to a lack of patch-level relevance modeling, noise-invariant embeddings, and class-anchored regularization. By unifying these mechanisms in a single QNN pipeline, QAPNet achieves significantly improved robustness and classification accuracy across modalities. Importantly, our noise model targets input image corruption, a realistic scenario in clinical imaging where degraded visuals coexist with bag-level annotation but uncorrupted labels, contrasting with typical label-noise assumptions. We evaluate QAPNet on seven datasets—including OASIS (MRI)

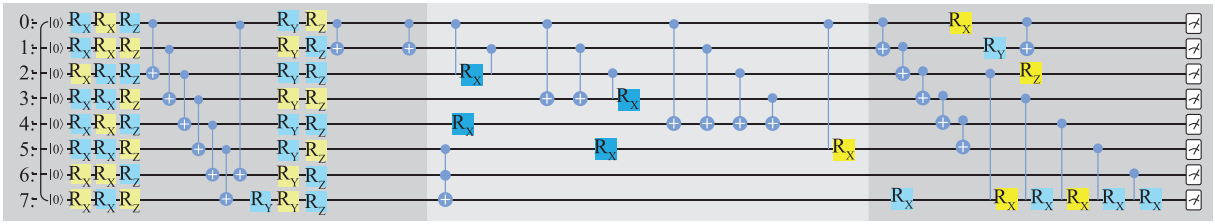

Figure 1: Structure of the 8-qubit VQC used in QAPNet: angle embedding on each qubit followed by $L = 5$ blocks of parameterized single-qubit rotations (R_X, R_Y, R_Z) interleaved with a ring entangler of standard CNOTs (filled dot = control, $\oplus$ = target); no non-standard gates are used.

and six MedMNIST v2 tasks (Yang, Shi et al. 2023)—under different Gaussian noise, consistently outperforming eight re-implemented baselines (AMIL, PQNet, SC-MIL, QCNN, QTNet, HQNet, QCCNet, A-HQCNN) in both accuracy and robustness. Our key contributions are as follows.

- We propose **QAPNet**, a hybrid quantum-classical framework that unifies sliding-window patch extraction, variational quantum encoding (without classical CNNs), element-wise attention, contrastive learning, and prototype anchoring for robust medical image classification.
- We introduce a new instance aggregation mechanism that is **element-wise, patch-independent**, and interpretable, yielding stronger generalization than dot-product or cross-instance attention under noise.
- We validate QAPNet on **seven noisy imaging datasets**, consistently outperforming classical and quantum baselines across modalities and Gaussian noise levels.

QAPNet addresses a fundamental gap in quantum learning by explicitly modeling noisy inputs and bag-level annotation, offering a scalable and interpretable solution for quantum-enhanced medical vision systems.

Proposed Methodology

We propose **QAPNet** (Quantum-Attentive Patch-based Network), a hybrid quantum-classical framework designed to perform medical image classification under noisy and bag-level annotation conditions. The **core intuition** behind QAPNet is to decompose image-level prediction into three levels of modeling: (i) instance-level quantum patch encoding using a variational quantum circuit (VQC), (ii) element-wise reweighting through attention to handle ambiguity in patch relevance, and (iii) embedding space alignment via contrastive learning and class-specific prototype regularization. Figure 2 presents the overall QAPNet process. An input image is partitioned into overlapping patches, each processed by an 8-qubit VQC (Figure 1). Embeddings are reweighted by additive attention and aggregated into a bag representation; we employ a VQC (rather than a classical MLP) to obtain parameter-efficient entangled features that empirically improve robustness under corruption, as quantified by qubit ablations in Table 3 with classical blocks held fixed.

Patch Extraction and MIL Formulation. Given a grayscale image $X \in \mathbb{R}^{H_{\text{img}} \times W_{\text{img}}}$ with a weak image-level

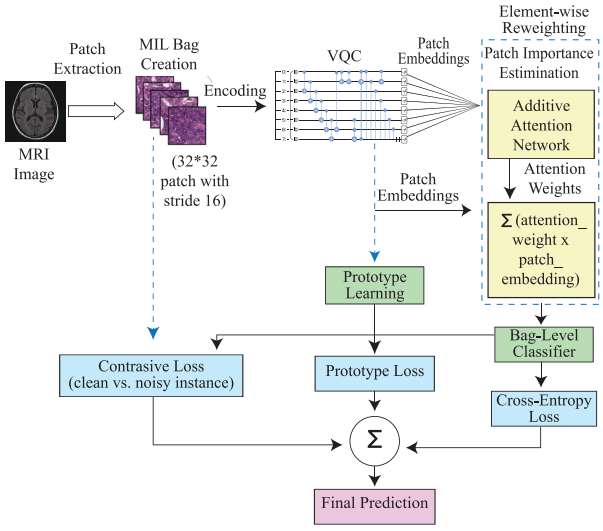

Figure 2: Overview of QAPNet architecture. Sliding window patch extraction forms MIL bags. Each patch is quantum-encoded via an 8-qubit variational circuit, then reweighted using element-wise additive attention, which is aggregated and supervised via classification, contrastive, and prototype losses.

$y \in \{1, \ldots, C\}$, we partition it into overlapping patches $\{x_i\}_{i=1}^{N}$ using a sliding window of size $s \times s$ and stride δ, where H_{img} and W_{img} denote the image height and width. Each patch x_i is treated as an instance in a Multiple Instance Learning (MIL) bag and inherits the image-level y. Unlike pre-selected region methods (e.g., HQNet, PQNet), our dense and overlapping extraction strategy ensures full spatial coverage and redundancy—critical for handling input-level corruption.

Quantum Patch Encoding with N-Qubit VQC. Each patch x_i is transformed into a latent embedding $h_i \in \mathbb{R}^d$ using a shared VQC. The process includes:

- **AngleEmbedding:** x_i is first uniformly pooled to an 8-dimensional vector $v_i = P, \text{vec}(x_i)$, then normalized and angle-encoded by applying $R_Y(v_{i,j})$ on each of the $N = 8$ qubits; no feature re-uploading is used.
- **Entanglement and Rotation:** The circuit applies L layers of a ring-topology CNOT entangler interleaved

with parameterized rotations R_X, R_Y, R_Z; in circuit diagrams, the CNOT is a filled dot and the target is $\oplus$.

- **Measurement:** The expectation values of Pauli-Z observables yield classical output:

$$h_{ik} = \langle \psi_i | Z_k | \psi_i \rangle, \quad k = 1, \ldots, d. \tag{1}$$

Element-wise Attention-Based Reweighting. To effectively aggregate instance-level embeddings into a noise-resilient global representation, QAPNet employs an *additive attention mechanism* that computes patch-level importance scores in a permutation-invariant and content-aware manner. Given a set of quantum-encoded patch embeddings $\{h_i\}_{i=1}^N$, each embedding $h_i \in \mathbb{R}^d$ is passed through a two-layer MLP with a non-linear activation to generate an unnormalized attention score:

$$e_i = \mathbf{u}^\top \tanh(W h_i + b), \tag{2}$$

where $W \in \mathbb{R}^{m \times d}$, $\mathbf{u} \in \mathbb{R}^m$, and $b \in \mathbb{R}^m$ are trainable parameters. These logits are then normalized using the softmax function to obtain the instance attention weights:

$$a_i = \frac{\exp(e_i)}{\sum_{j=1}^N \exp(e_j)}. \tag{3}$$

The final bag-level embedding $z \in \mathbb{R}^d$ is computed as a weighted sum of the patch embeddings:

$$z = \sum_{i=1}^N a_i h_i. \tag{4}$$

This element-wise reweighting mechanism allows QAPNet to dynamically prioritize diagnostically relevant patches while suppressing the influence of noisy or non-informative regions. Unlike dot-product attention mechanisms used in SC-MIL or QCCNet, which model inter-instance interactions and may amplify interference among noisy patches, the additive attention design ensures that each patch is scored independently for improving interpretability and noise tolerance.

Noise-Invariant Learning via Contrastive and Prototype Regularization. To further enhance robustness under noisy input conditions, QAPNet introduces two complementary regularization strategies: contrastive alignment and prototype-guided clustering. These techniques are especially effective in medical imaging, where Gaussian corruption of intensity and anatomy-like variability often causes models to overfit superficial or noise-driven cues.

Contrastive Learning for Embedding Invariance. We adopt a supervised contrastive formulation to enforce that noisy images yield similar latent representations to their clean counterparts. For a training image X and its noisy augmented version X' (obtained by applying Gaussian noise with $\sigma = 0.3$), we extract respective embeddings $z = f(X)$ and $z' = f(X')$ using the same QAPNet pipeline. To prevent the model from learning noise-specific distortions, we apply the following contrastive loss, adapted from SimCLR and NT-Xent:

$$L_{\text{con}} = -\log \frac{\exp(\text{sim}(z, z')/\tau)}{\sum_{j=1}^B \exp(\text{sim}(z, z'_j)/\tau)}, \tag{5}$$

where $\text{sim}(z, z') = \frac{z^\top z'}{\|z\| \|z'\|}$ is the cosine similarity between normalized embeddings, τ is a temperature scaling parameter, and B is the batch size. This loss encourages representations of the same image under noise (z, z') to remain close while separating different samples in the batch. Unlike QTNet or PQNet, which lack any form of noise-aware regularization where QAPNet explicitly penalizes latent drift caused by corruption and promoting generalization across noisy imaging modalities.

Prototype-Based Semantic Anchoring. To structure the latent space and stabilize learning under bag-level annotation, we introduce a class-prototype alignment objective. Specifically, we maintain a set of trainable class prototypes $\{c_k\}_{k=1}^C \subset \mathbb{R}^8$, one for each class. These prototypes act as semantic centroids and we minimize the squared Euclidean distance between the bag-level embedding z and the prototype c_y corresponding to its label:

$$L_{\text{proto}} = \|z - c_y\|_2^2. \tag{6}$$

Although bag-level annotation, the image-level label y provides sufficient class guidance at the bag level to enable prototype-based regularization to anchor noisy embeddings to semantic centroids—mitigating feature drift observed in models like A-HQCNN and QCNN and improving both intra-class cohesion and inter-class separability.

Algorithm 1 presents a detailed forward pass of QAPNet, integrating all components from patch-level encoding to multi-loss optimization. This routine is designed to optimize quantum parameters Θ_Q, attention weights, and prototype vectors in a cohesive loop. QAPNet addresses the challenges of noisy medical imaging by combining quantum-enhanced patch encoding, element-wise reweighting, and prototype-guided contrastive learning. It enforces noise-invariant, semantically structured embeddings—bridging key gaps in prior methods and delivering robust and interpretable performance across all evaluated noise settings.

Competing Methods

We benchmark QAPNet against eight re-implemented baseline models, spanning classical, hybrid, and quantum paradigms. All baselines were evaluated under identical datasets, patching strategies, and Gaussian image noise settings for fair comparison.

- **AMIL** (Ilse, Tomczak, and Welling 2018b) uses an additive attention mechanism for MIL bag aggregation but assumes clean bag-level labels. Its lack of noise filtering degrades robustness under image corruption, unlike QAPNet's instance-aware reweighting.
- **PQNet** (Mahmud and Fattah 2024) employs a shallow quantum classifier without contrastive or MIL structure. It underperforms in multi-class, noisy settings due to limited scalability and a lack of prototype supervision.

Dataset	Noise	AMIL	PQNet	SC-MIL	QCNN	QTNet	HQNet	QCCNet	A-HQCNN	QAPNet
OASIS	5%	78.46	77.21	81.23	61.00	37.85	35.07	38.50	36.89	**84.86**
	10%	63.56	59.50	67.32	45.88	24.50	21.50	27.03	25.53	**82.16**
	30%	56.87	47.10	59.10	33.05	16.37	19.23	16.58	20.70	**79.87**
PathMNIST	5%	61.85	63.52	70.32	37.55	22.78	31.67	21.66	12.78	**72.16**
	10%	53.12	54.32	61.55	35.22	16.56	20.85	14.33	10.56	**71.98**
	30%	43.95	49.22	53.59	30.55	10.00	10.56	7.44	10.00	**71.26**
BloodMNIST	5%	48.84	51.23	52.53	43.74	20.62	21.25	19.37	19.00	**52.90**
	10%	37.39	43.52	41.20	38.12	17.50	18.88	18.12	16.25	**51.92**
	30%	31.78	37.25	32.50	29.37	12.25	16.25	10.62	13.75	**50.16**
OrganAMNIST	5%	54.43	59.25	64.36	47.27	21.36	21.82	18.18	19.55	**65.08**
	10%	51.49	58.96	53.23	44.54	18.18	16.82	7.27	16.82	**64.77**
	30%	47.33	51.58	44.35	38.18	11.55	15.45	6.36	11.82	**61.55**
TissueMNIST	5%	53.89	56.23	60.00	48.12	23.12	15.62	20.87	23.75	**63.90**
	10%	42.86	48.36	49.36	37.50	11.25	14.38	18.12	11.25	**62.29**
	30%	35.52	42.85	40.25	31.25	6.25	11.25	9.38	9.38	**60.28**
OCTMNIST	5%	48.23	51.30	**53.57**	30.00	28.75	35.00	36.25	30.00	52.28
	10%	39.56	42.20	43.29	21.00	27.50	30.00	21.25	23.25	**51.39**
	30%	33.28	38.32	35.30	15.25	21.25	28.75	17.00	16.25	**49.14**
DermaMNIST	5%	48.52	52.56	55.36	50.00	20.00	26.43	17.14	24.29	**56.36**
	10%	43.35	46.25	46.06	39.28	12.86	19.43	11.42	19.71	**54.45**
	30%	31.95	37.58	36.28	35.71	8.57	18.57	9.42	14.29	**52.71**

Table 1: Classification accuracy (%) of QAPNet and eight baseline models across seven medical imaging datasets under three levels of Gaussian input noise (5%, 10%, and 30%). QAPNet consistently achieves the highest accuracy across all settings, demonstrating superior robustness to noise perturbations and generalization across diverse domains. Bold values indicate the best performance per row.

- **SC-MIL** (Ju et al. 2022a) introduces attention refinement under MIL but depends on CNN encoders and lacks embedding regularization. This leads to overfitting under noise, which QAPNet addresses via quantum compression and contrastive learning.
- **QCNN** (Bokhan et al. 2022) implements quantum convolution and pooling for clean image classification. However, it lacks any mechanisms to cope with noisy or corrupted inputs.
- **QTNet** (Dhara, Agrawal, and Roy 2024) combines classical ResNet18 features with quantum classification. While effective on clean data, it fails to correct or suppress corrupted patches or noisy labels.
- **HQNet** (Mazher et al. 2024a) fuses CNN features with quantum layers but suffers under label noise due to reliance on classical encoders and absence of patch-level modeling.
- **QCCNet** (Yousif, Al-Khateeb, and Garcia-Zapirain 2024c) is optimized for efficient quantum convolution, yet lacks attention or robustness regularization, resulting in poor performance under input degradation.
- **A-HQCNN** (Ajlouni et al. 2023a) augments HQCNN with adaptive optimization. Although it improves training stability, it fails to handle noisy instances or enforce latent space structure.

Datasets, Hyperparameter Settings, and Experimental Setup

Datasets. We evaluate QAPNet on seven benchmark medical imaging datasets. Six subsets from MedMNISTv2 (Yang, Shi et al. 2023) and a curated version of the OASIS-1 MRI dataset (Daithal 2022). These cover diverse imaging modalities (MRI, CT, OCT, dermoscopy, microscopy) and diagnostic challenges. To simulate real-world image degradation, we inject Gaussian noise into the image content at 5%, 10%, and 30% power levels, modeling acquisition variability such as scanner artifacts and resolution loss—distinct from label noise. Unlike asymmetric label noise, which often stems from human error or inter-class confusion (e.g., visually similar diagnoses), our focus is on robust representation under corrupted inputs. We do not use noisy-label datasets like CheXpert, MIMIC-CXR, or Camelyon due to their massive size, limited MIL compatibility, or incompatibility with quantum simulation constraints. Instead, we select datasets enabling efficient simulation (via PennyLane), patch-based learning, and controlled noise benchmarking. Specifically, **OASIS** comprises 80,000+ axial MRI slices from 461 patients (4 CDR-based dementia stages); **PathMNIST** features 9-class colorectal histology images; **BloodMNIST** offers 8-class blood smear cells; **OrganAMNIST** and **OrganCMNIST** include CT and contrast-enhanced abdominal scans (11 classes); **OCTMNIST** contains 4-class retinal OCT data; and **DermaMNIST** provides 7-class dermatoscopic lesion images. Together, these benchmarks offer a rigorous yet tractable foundation for evaluating QAPNet's quantum patch encoding, attention-based instance reweighting, and prototype-guided contrastive learning under structured image corruption.

Hyperparameter Settings and Experimental Setup. All models—including QAPNet and the eight baselines—were

Algorithm 1: Forward Pass of QAPNet

Input: Image X, label y, QNN parameters Θ_Q, prototypes $\{c_k\}$.
Output: Prediction $\hat{y}$, loss $\mathcal{L}_{\text{total}}$.

Extract overlapping patches $\{x_i\}_{i=1}^N$ from X using 32×32 window and stride 16.
for each x_i **do**
  Apply AngleEmbedding and $U(\Theta_Q)$ to encode quantum state:
  $|\psi_i\rangle = U(x_i; \Theta_Q)|0\rangle^{\otimes N}$
  Measure Pauli-Z expectations:
  $h_i = [\langle\psi_i|Z_k|\psi_i\rangle]_{k=1}^N \in \mathbb{R}^N$
end for
Compute attention logits: $e_i = \mathbf{u}^\top \tanh(Wh_i + b)$
Normalize with softmax: $a_i = \frac{\exp(e_i)}{\sum_{j=1}^N \exp(e_j)}$
Compute bag embedding: $z = \sum_{i=1}^N a_i h_i$
Augment X to X' with Gaussian noise; repeat steps to obtain z'
Compute classifier output: $\hat{p} = \text{softmax}(W_o z + b_o)$
Compute losses:

$$\mathcal{L}_{\text{cls}} = -\log \hat{p}_y$$

$$\mathcal{L}_{\text{con}} = -\log \frac{\exp(\text{sim}(z, z')/\tau)}{\sum_j \exp(\text{sim}(z, z'_j)/\tau)}$$

$$\mathcal{L}_{\text{proto}} = \|z - c_y\|_2^2$$

Aggregate total loss:
$\mathcal{L}_{\text{total}} = \lambda_{\text{cls}} \cdot \mathcal{L}_{\text{cls}} + \lambda_{\text{con}} \cdot \mathcal{L}_{\text{con}} + \lambda_{\text{proto}} \cdot \mathcal{L}_{\text{proto}}$
return $\hat{y} = \arg\max_k \hat{p}_k$, $\mathcal{L}_{\text{total}}$

trained using the Adam optimizer (learning rate 0.001, batch size 128) for up to 30 epochs with early stopping on validation loss. The loss function combines cross-entropy ($\mathcal{L}_{\text{cls}}$), supervised contrastive loss ($\mathcal{L}_{\text{con}}$) with temperature $\tau = 0.5$, and prototype regularization ($\mathcal{L}_{\text{proto}}$), balanced using $\lambda_{\text{con}} = \lambda_{\text{proto}} = 1.0$. Accuracy (ACC) is used as the primary evaluation metric. Training was conducted using the Pennylane simulator with GPU acceleration (Ubuntu 20.04, Python 3.9, TensorFlow 2.6, PyTorch 1.10, PennyLane 0.25). Input images were divided into ∼40–50 overlapping patches via a 32×32 sliding window (stride 16), used as MIL instances. Each patch was encoded using a shared 8-qubit variational quantum circuit (VQC) comprising: (i) *AngleEmbedding* via R_Y rotations, (ii) parameterized R_X, R_Y, R_Z gates, (iii) circular entanglement with CNOT gates, and (iv) Pauli-Z expectation measurement to produce an 8-dimensional embedding. We use $L = 5$ entangling layers to balance expressive depth with NISQ-era feasibility. All network components—including Θ_Q, additive attention weights, and class prototypes—are optimized jointly using backpropagation and parameter-shift rules.

Results and Discussion

We evaluate QAPNet across multiple noisy and bag-level annotation medical imaging benchmarks for validating the impact of each architectural component on classification robustness and generalization. The analysis includes baseline comparisons and ablation studies on instance aggregation, quantum encoding, and patch selection.

Dataset	Noise	Elem. Reweight.	Cross-Inst.
OASIS	5%	**84.86**	84.60
	10%	**82.16**	73.16
	30%	**79.87**	66.00
PathMNIST	5%	**72.16**	67.55
	10%	**71.98**	55.22
	30%	**71.26**	50.55
BloodMNIST	5%	52.90	**53.74**
	10%	**51.92**	39.32
	30%	**50.16**	28.39
OrganAMNIST	5%	**65.08**	57.27
	10%	**64.77**	45.44
	30%	**61.55**	38.88
TissueMNIST	5%	**63.90**	58.22
	10%	**62.29**	47.57
	30%	**60.28**	36.55
OCTMNIST	5%	**52.28**	50.00
	10%	**51.39**	41.00
	30%	**49.14**	35.28
DermaMNIST	5%	**56.36**	53.70
	10%	**54.45**	39.26
	30%	**52.71**	35.85

Table 2: Ablation study comparing **element-wise instance reweighting** (our method) and **cross-instance interaction** (dot-product-based aggregation) across seven datasets and three Gaussian noise levels. Element-wise reweighting consistently outperforms under higher noise. Bolded values denote better accuracy.

Performance Comparison with Baseline Models.

We benchmark **QAPNet** against eight strong baselines across seven medical datasets under 5%, 10%, and 30% Gaussian noise. As summarized in Table 1, QAPNet consistently achieves the highest accuracy across all datasets and noise levels, demonstrating robust generalization under image inputs. For example, on **OASIS**, QAPNet yields 84.86%, 82.16%, and 79.87% accuracy at 5%, 10%, and 30% noise, respectively—outperforming SC-MIL by up to 20%. Similar resilience is observed on **PathMNIST** (71.26% at 30% noise) compared to sharp degradation in SC-MIL (53.59%), QTNet (10%), and QCCNet (7.44%). These results highlight QAPNet's architectural advantages: quantum patch encoding eliminates dependence on classical CNNs, additive attention offers localized interpretability and contrastive-prototype learning promotes semantic structure and noise invariance. In contrast, baselines lacking these mechanisms suffer significant drops under heavy noise.

Ablation Study I: Instance Aggregation Mechanisms.

We assess the impact of our element-wise reweighting mechanism by comparing it with cross-instance dot-product aggregation across all datasets and noise levels (Table 2). QAPNet's instance-independent attention consistently outperforms cross-instance interaction, especially under high noise. For example, at 30% noise, it achieves 79.87% on **OASIS** (vs. 66.00%), 71.26% on **PathMNIST** (vs. 50.55%), and 61.55% on **OrganAMNIST** (vs. 38.88%). These results validate that localized, per-instance weighting is more effective for bag-level annotations and noisy medical imaging tasks, offering both interpretability and resilience over con-

Dataset Size	Patch Selection	Qubits	Accuracy (%)
1952 (488 each)	Dynamic (5)	4	32.13
1952 (488 each)	Dynamic (5)	8	43.15
Full (86400)	Dynamic (5)	4	62.33
Full (86400)	Dynamic (5)	8	67.12
1952 (488 each)	Dynamic (15)	4	46.58
1952 (488 each)	Dynamic (15)	8	53.00
Full (86400)	Dynamic (15)	4	75.35
Full (86400)	Dynamic (15)	8	79.23
1952 (488 each)	Sliding Window	4	48.62
1952 (488 each)	Sliding Window	8	68.62
Full (86400)	Sliding Window	4	78.89
Full (86400)	Sliding Window	8	**84.89**

Table 3: Ablation study evaluating patch selection strategies (*dynamic* vs. *sliding window*) and quantum encoding capacity (4 vs. 8 qubits) across different dataset sizes. Sliding window with 8 qubits on the full dataset achieves the best accuracy, validating our design choice in QAPNet.

ventional dot-product MIL.

Ablation Study II: Patch Strategy and Quantum Circuit Depth.

We analyze the influence of patch extraction and quantum encoder capacity (Table 3). Holding the attention MLP and classifier fixed, increasing qubits from 4 to 8 (a linear increase in $3NL$ parameters) consistently improves accuracy, isolating the VQC's contribution to the end-to-end gains. Moreover, our **sliding window** strategy outperforms random sampling consistently—e.g., achieving 85.00% accuracy (vs. 81.08%) on the full dataset with 8 qubits—due to systematic coverage of pathology-rich regions.

Evaluation and Analysis

We conduct a comprehensive evaluation of QAPNet to validate its robustness, interpretability, and hyperparameter sensitivity under noisy supervision. Our evaluation covers three dimensions: attention-based interpretability, latent space consistency and sensitivity analysis.

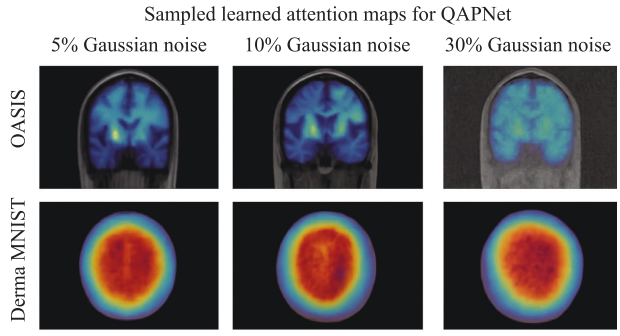

Figure 3: Sample learned attention maps for QAPNet on OASIS and DermaMNIST at 5%, 10%, and 30% Gaussian noise levels. Attention localizes informative regions while reducing sensitivity to noisy areas.

Visualizing Attention-Based Reweighting. To demonstrate the effectiveness of our element-wise attention mechanism, we visualize learned attention maps on representative samples from OASIS (MRI) and DermaMNIST (dermatology) under 5%, 10%, and 30% Gaussian noise. As shown in Figure 3, QAPNet accurately emphasizes disease-relevant regions (e.g., cortical boundaries, lesion cores) while de-emphasizing corrupted or non-discriminative areas. As noise severity increases, attention becomes more localized and noise-adaptive, confirming our model's capability to focus on robust visual cues.

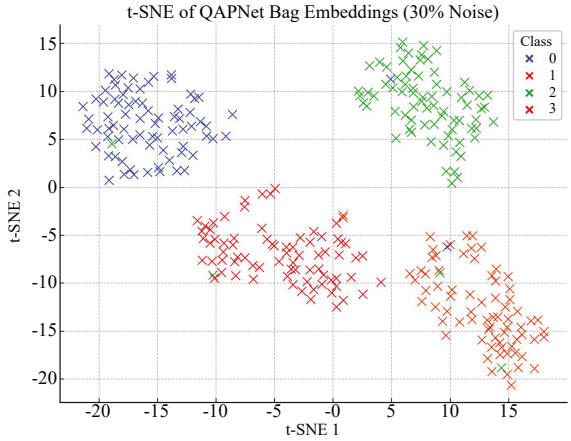

Figure 4: t-SNE projection of bag-level embeddings on OASIS (30% noise). QAPNet preserves separable, class-consistent latent clusters despite input degradation.

Latent Space Consistency under Noise.

We evaluate the semantic stability of QAPNet's embedding space using t-SNE visualization of bag-level embeddings from OASIS at 30% noise. Figure 4 shows that class-wise clusters remain well-separated despite corruption, suggesting that our contrastive and prototype-based losses promote noise-invariant, semantically aligned representations.

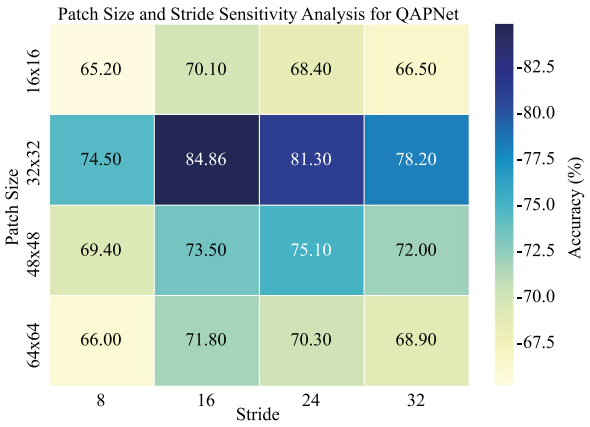

Figure 5: Patch size and stride sensitivity. QAPNet performs best at 32×32 patch size and stride 16, offering an optimal trade-off between spatial detail and context.

Hyperparameter Sensitivity.

We analyze the sensitivity of hyperparameters as follows. **(i) Patch and Stride Analysis.** We assess the sensitivity of QAPNet to patch size and stride in Figure 5. Among all tested configurations, the 32×32 patch size with stride 16 achieves the highest accuracy (84.86%), likely due to its balance of local detail and coverage. Smaller patches lack context, while larger patches dilute localization ability. **(ii) Loss Balancing Factors.** We analyze the joint effect of loss weights λ_{proto} and λ_{con} in Figure 6. The best result (84.8%) is achieved at $(\lambda_{\text{proto}}, \lambda_{\text{con}}) = (1.0, 1.0)$. This confirms that both prototype regularization and contrastive consistency are essential, and overly aggressive regularization can impair performance.

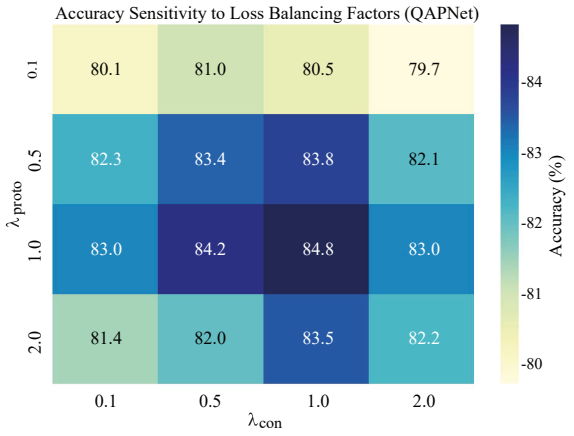

Figure 6: Sensitivity to loss balancing factors. Moderate regularization $(\lambda_{\text{proto}}, \lambda_{\text{con}}) = (1.0, 1.0)$ yields optimal performance.

Related Work

Quantum Learning for Noisy Medical Data. Recent progress in quantum neural networks (QNNs) has demonstrated their potential for medical imaging (Ajlouni et al. 2023b; Yousif, Al-Khateeb, and Garcia-Zapirain 2024b; Mazher et al. 2024b; Mathur et al. 2021). However, these approaches are primarily validated on clean datasets and often rely on classical backbones. In contrast, QAPNet directly encodes patches using variational quantum circuits and is explicitly optimized for noise robustness via contrastive and prototype objectives.

Robust Learning Under Label Noise. Robust classification under label noise has been addressed using sample reweighting (Shu et al. 2019), loss correction (Damian et al. 2019; Rahman and Zhuang 2026), and dual-uncertainty strategies (Ju et al. 2022b). In medical imaging, weak supervision and ambiguous labels are particularly common (Khanal, Shrestha, and Linte 2023; Xue et al. 2022b). Unlike these classical methods, QAPNet uses quantum patch representations and attention-based instance aggregation to suppress noisy regions, while enforcing semantic structure via prototypes.

Contrastive and Prototype Regularization. Contrastive learning enhances noise invariance by aligning views of the same sample, while prototype-based losses encourage intra-class compactness (Fostiropoulos and Itti 2022; Snell, Swersky, and Zemel 2017). QAPNet integrates both into a hybrid quantum architecture, enabling interpretable and structured learning in noisy latent spaces—an approach largely unexplored in quantum machine learning (QML).

Attention-Based MIL for Patch Aggregation. Multiple Instance Learning (MIL) has been applied for bag-level annotations of medical data, with attention-based pooling allowing instance importance estimation (Ilse, Tomczak, and Welling 2018a). Recent works like SC-MIL (Juyal et al. 2024) improve robustness via consistency regularization. Our model builds upon additive attention but replaces classical encoders with a quantum feature extractor and integrates latent regularization losses, extending MIL to the quantum domain.

Limitations and Future Work

QAPNet is evaluated in classical simulation using PennyLane's `default.qubit` (i.e., without hardware noise). We will release code with exact seeds, configs, preprocessing, and baseline settings for full reproducibility. We report single-seed accuracy without error bars due to limited space; multi-seed means±std and runtime/memory profiles will appear in the repository. Our study focuses on additive Gaussian noise, extending to other corruptions and to ablations that replace the VQC with MLP/CNN feature extractors, as well as per-loss ablations, which are planned in the public release. The shared patch-level circuit is fixed for efficiency, where learning task-adaptive circuits and deploying on real hardware are promising directions. In the future, QAPNet may be implemented in real quantum hardware and be extended to segmentation and multi-label classification with deeper VQCs.

Conclusion

In this work, we proposed **QAPNet**, a quantum-attentive patch-level network for robust medical image classification under noisy bag-level annotations. QAPNet combines variational quantum encoding, element-wise attention, and contrastive-prototype regularization into a unified, interpretable pipeline that does not rely on classical CNNs. Across seven medical imaging benchmarks with varying levels of Gaussian corruption, QAPNet consistently outperforms eight strong baselines, and ablations confirm the importance of sliding-window patch selection, quantum feature extraction, and noise-aware instance reweighting. The model generalizes across diverse modalities, indicating its suitability for real-world deployment and advancing the frontier of QML for medical imaging. Beyond these empirical gains, our results suggest that quantum-enhanced patch representations can serve as a general backbone for robust perception in high-noise healthcare settings, and that principled noise-aware representation learning will be key to building trustworthy quantum-classical decision support systems in clinical workflows.

Acknowledgments

This work was supported in part by the National Science Foundation under Grant No. 2451670.

References

Ajlouni, N.; Özyavaş, A.; Takaoğlu, M.; Takaoğlu, F.; and Ajlouni, F. 2023a. Medical image diagnosis based on adaptive hybrid quantum CNN. *BMC Medical Imaging*, 23(1): 126.

Ajlouni, N.; et al. 2023b. Medical image diagnosis based on adaptive hybrid quantum cnn. *BMC Medical Imaging*, 23(1): 126.

Bokhan, D.; Mastiukova, A. S.; Boev, A. S.; Trubnikov, D. N.; and Fedorov, A. K. 2022. Multiclass classification using quantum convolutional neural networks with hybrid quantum-classical learning. *Frontiers in Physics*, 10: 1069985.

Cunningham, J.; and Zhuang, J. 2025. Investigating and mitigating barren plateaus in variational quantum circuits: a survey. *Quantum Information Processing*, 24: 48.

Daithal, N. 2022. Images - OASIS Brain MRI Classification. https://www.kaggle.com/datasets/ninadaithal/imagesoasis. Accessed: 2025-07-26.

Damian, A.; Warmuth, M. K.; Zhang, L.; Chen, B.; and Li, F. 2019. Analysis of classifiers robust to noisy labels. In *International Conference on Machine Learning (ICML)*, volume 97, 1500–1509.

Dhara, B.; Agrawal, M.; and Roy, S. 2024. Multi-class classification using quantum transfer learning. *Quantum Information Processing*, 23(2): 34.

Fostiropoulos, I.; and Itti, L. 2022. Supervised contrastive prototype learning: Augmentation-free robust neural network. *arXiv preprint arXiv:2211.14424*.

Ilse, M.; Tomczak, J.; and Welling, M. 2018a. Attention-based Deep Multiple Instance Learning. In *International Conference on Machine Learning*, 2127–2136.

Ilse, M.; Tomczak, J. M.; and Welling, M. 2018b. Attention-based deep multiple instance learning. In *International conference on machine learning*, 2127–2136. PMLR.

Ju, L.; Wang, X.; Wang, L.; Mahapatra, D.; Zhao, X.; Zhou, Q.; Liu, T.; and Ge, Z. 2022a. Improving medical images classification with label noise using dual-uncertainty estimation. *IEEE Transactions on Medical Imaging*, 41(6): 1533–1546.

Ju, L.; et al. 2022b. Improving medical images classification with label noise using dual-uncertainty estimation. *IEEE Transactions on Medical Imaging*, 41(6): 1533–1546.

Juyal, R.; Chen, C.-H.; Kalra, S.; and Martel, A. 2024. SC-MIL: Supervised Contrastive Multiple Instance Learning for Pathology Image Classification. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2001–2010.

Karimi, D.; Dou, H.; Warfield, S.; and Gholipour, A. 2020. Deep learning with noisy labels: Exploring techniques and remedies in medical image analysis. *Medical image analysis*, 65: 101759.

Khanal, B.; Shrestha, P.; and Linte, C. A. 2023. Improving medical image classification in noisy labels using only self-supervised pretraining. In *MICCAI Workshop on Data Engineering in Medical Imaging (DEMI)*.

Liu, J.; Lim, K. H.; Wood, K. L.; et al. 2021. Hybrid quantum-classical convolutional neural networks. *Science China Physics, Mechanics & Astronomy*, 64(9): 290311.

Mahmud, J.; and Fattah, S. A. 2024. Patch-Based End-to-End Quantum Learning Network for Reduction and Classification of Classical Data. ArXiv preprint, arXiv:2409.15214.

Mathur, N.; et al. 2021. Medical image classification via quantum neural networks. *arXiv preprint arXiv:2109.01831*.

Mazher, M.; Qayyum, A.; Khan, M. A.; Niederer, S.; Mokayef, M.; and Hassan, C. 2024a. Hybrid classical and quantum deep learning models for medical image classification. *In Proc. of ICAROB*, 222–226.

Mazher, M.; et al. 2024b. Hybrid classical and quantum deep learning models for medical image classification. *ICAROB*.

Rahman, M.; and Zhuang, J. 2026. NQNN: Noise-Aware Quantum Neural Networks for Medical Image Classification. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2025*, 433–442.

Shu, J.; Xie, Q.; Yi, L.; Zhao, Q.; Zhou, S.; Xu, Z.; and Meng, D. 2019. Meta-weight-net: Learning an explicit mapping for sample weighting. In *Advances in Neural Information Processing Systems*, volume 32.

Snell, J.; Swersky, K.; and Zemel, R. S. 2017. Prototypical networks for few-shot learning. In *Proceedings of the 31st International Conference on Neural Information Processing Systems (NeurIPS)*, 4077–4087.

Xue, C.; Yu, L.; Chen, P.; Dou, Q.; and Heng, P.-A. 2022a. Robust medical image classification from noisy labeled data with global and local representation guided co-training. *IEEE transactions on medical imaging*, 41(6): 1371–1382.

Xue, C.; et al. 2022b. Robust medical image classification from noisy labeled data with global and local representation guided co-training. *IEEE Transactions on Medical Imaging*, 41(6): 1371–1382.

Yang, J.; Shi, R.; et al. 2023. MedMNIST v2 – A large-scale lightweight benchmark for 2D and 3D biomedical image classification. *Scientific Data*, 10(1): 41.

Yousif, M.; Al-Khateeb, B.; and Garcia-Zapirain, B. 2024a. A new quantum circuit of quantum convolutional neural network for X-ray images classification. *IEEE Access*, 12: 65658–65667.

Yousif, M.; Al-Khateeb, B.; and Garcia-Zapirain, B. 2024b. A new quantum circuit of quantum convolutional neural network for X-ray images classification. *IEEE Access*.

Yousif, M.; Al-Khateeb, B.; and Garcia-Zapirain, B. 2024c. A new quantum circuits of quantum convolutional neural network for x-ray images classification. *IEEE Access*.

Zhuang, J.; Cunningham, J.; and Guan, C. 2024. Improving trainability of variational quantum circuits via regularization strategies. *arXiv preprint arXiv:2405.01606*.

Zhuang, J.; and Guan, C. 2025. Large language models can help mitigate barren plateaus. *arXiv preprint arXiv:2502.13166*.